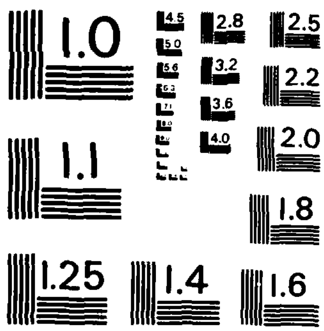
AD-A160 878 DEVELOPMENT OF A METHODOLOGY TO TEST VIDEO CODECS USED IN TELECONFERENCING(U) DELTA INFORMATION SYSTEMS INC HORSHAM PA 23 AUG 85 NCS-TIB-85-1 DCA100-83-C-0047 1/1

UNCLASSIFIED F/G 17/2 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

2

NCS TIB 85-1

NATIONAL COMMUNICATIONS SYSTEM

AD-A160 878

TECHNICAL INFORMATION BULLETIN 85-1

DEVELOPMENT OF A METHODOLOGY TO TEST VIDEO CODECS USED IN TELECONFERENCING

AUGUST 1985

SECURITY CLASSIFICATION OF THIS PAGE

AD-A160 878

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp. Date: Jun 30 1986

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Unlimited DISTRIBUTION STATEMENT A Approved for public release; Distribution Unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) NCS TIB 85-1
6a NAME OF PERFORMING ORGANIZATION Delta Information Systems, Inc. 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION National Communications System
6c. ADDRESS (City, State, and ZIP Code) Horsham Business Center, Bldg. 3 300 Welsh Road Horsham, PA 19044	7b. ADDRESS (City, State, and ZIP Code) ATTN: NCS-TS Washington, DC 20305-2010
8a. NAME OF FUNDING / SPONSORING ORGANIZATION National Communications System 8b OFFICE SYMBOL (If applicable) NCS-TS	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract DCA100-83-C-0047
8c. ADDRESS (City, State, and ZIP Code) Washington, D. C. 20305-2010 ATTN: NCS-TS	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 33127K PROJECT NO: TASK NO: WORK UNIT ACCESSION NO.:

11 TITLE (Include Security Classification)
Development of a Methodology to Test Video Codecs Used in Teleconferencing

12 PERSONAL AUTHOR(S)

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM ______ TO ______	85-08-23	51

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Video Teleconferencing, Test Methodology, Codecs

19 ABSTRACT (Continue on reverse if necessary and identify by block number) The purpose of this report is to define a test methodology for the evaluation of motion television codecs for teleconferencing applications. The results of this evaluation will serve as a guideline for the future preparation of specifications for codecs of this type. The principle characteristics of the codecs include the following: (1) utilize digital communication channels, (2) operate at a data rate of 1.544 Mbps, (3) Provide color capability, and (4) Provide motion capability. The specific objective of the tests described is to rank the codecs tested in order of performance capability. The reason that new tests are being developed is that there are presently no agreed upon test procedures for that purpose. The tests described will utilize a specially prepared video tape containing still and motion sequences designed specifically for the evaluation of this type of codec. These sequences will be passed through the codecs and the output recorded on video tape. The evaluation and grading of each codec will be on a subjective, comparative basis. The intent of CCIR Recommendation 500-2, Method for the subjective Assessment of the Quality of Television Pictures (Vol. XI, Part 1, XVth Plenary

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Dennis Bodson	22b TELEPHONE (Include Area Code) (202) 692-2124 22c OFFICE SYMBOL NCS-TS

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

NCS TECHNICAL INFORMATION BULLETIN 85-1

DEVELOPMENT OF A METHODOLOGY TO TEST VIDEO CODECS USED IN TELECONFERENCING

AUGUST 1985

PROJECT OFFICER

DENNIS BODSON
Senior Electronics Engineer
Office of NCS Technology
and Standards

APPROVED FOR PUBLICATION:

Marshall L. Cain

MARSHALL L. CAIN
Assistant Manager
Office of NCS Technology
and Standards

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunication Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunication Standards Committee identifies, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunication systems or to the achievement of a compatible and efficient interface between computer and telecommunication systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the Electronic Industries Association, the American National Standards Institute, the International Organization for Standardization, and the International Telegraph and Telephone Consultative Committee of the International Telecommunication Union. This Technical Information Bulletin presents an overview of an effort which is contributing to the development of compatible Federal, national, and international standards in the area of Video Teleconferencing. It has been prepared to inform interested Federal activities of the progress of these efforts. Any comments, inputs or statements of requirements which could assist in the advancement of this work are welcome and should be addressed to:

Office of the Manager
National Communications System
ATTN: NCS-TS
Washington, DC 20305
(202) 692-2124

DEVELOPMENT OF
A METHODOLOGY TO
TEST VIDEO CODECS
USED IN TELECONFERENCING

August 23, 1985

Final Report

Submitted to:

NATIONAL COMMUNICATIONS SYSTEM

Office of Technology and Standards

Washington, DC 20305

Contracting Agency:

DEFENSE COMMUNICATIONS AGENCY

Contract Number - DCA100-83-C-0047

Modification - P00004

Task - 3

Subtask - 2

D E L T A I N F O R M A T I O N S Y S T E M S, I N C.
Horsham Business Center, Bldg. 3
300 Welsh Road
Horsham PA 19044

PAGE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

SECTION 1 INTRODUCTION AND SUMMARY

This document summarizes work performed by Delta Information Systems, Inc. for the Office of Technology and Standards of the National Communications Systems, an organization of the U.S. Government, under Contract Number DCA100-83-C-0047 Modification P00004. The work was performed under Subtask 2 (Development of Test Methodology) under Task 3. The Office of Technology and Standards, headed by National Communications System Assistant Manager Marshall L. Cain, is responsible for the management of the Federal Telecommunications Standards Program, which develops telecommunication standards whose use is mandatory by all Federal agencies.

The purpose of this report is to define a test methodology for the evaluation of motion television codecs for teleconferencing applications. The results of this evaluation will serve as a guideline for the future preparation of specifications for codecs of this type.

The principal characteristics of the codecs include the following:

- Utilize digital communication channels
- Operate at a data rate of 1.544 Mbps
- Provide color capability
- Provide motion capability

The specific objective of the tests described is to rank the codecs tested in order of performance capability. The reason that

new tests are being developed is that there are presently no agreed upon test procedures for that purpose. The tests described will utilize a specially prepared video tape containing still and motion sequences designed specifically for the evaluation of this type of codec. These sequences will be passed through the codecs and the output recorded on video tape. The evaluation and grading of each codec will be on a subjective, comparative basis. The intent of CCIR Recommendation 500-2, Method for the subjective Assessment of the Quality of Television Pictures (Vol. XI, Part 1, XVth Plenery Assembly, Geneva 1982) will serve as a guideline. In addition, a number of video test signals will be passed through the codecs and recorded for future objective evaluation and correlation with the subjective test results.

Section 2 outlines the general test philosophy and the steps that will have to be followed in the performance of the tests. Section 3 discusses the parameters to be tested, both subjectively and objectively, and includes a list of desirable test signals. Section 4 briefly describes the procedure and steps needed to gather data on the codecs under test. The main output of this test methodology development study is contained in Section 5 which covers the evaluation of the codec output test tapes. Several possible methods are discussed and a preferred procedure is recommended. Section 6 gives a conclusion and recommendations for related further efforts. Details of a concept for objective motion evaluation are contained in Appendix A.

SECTION 2 TEST PHILOSOPHY

Testing and evaluating digital television codecs capable of conveying motion in a comparatively narrow data channel presents new problems in testing concepts. This is because there are no standardized tests for this purpose. Furthermore, in an actual teleconferencing application, the resultant picture as received and displayed by the codec is evaluated by the viewer, consciously or otherwise, against what he is familiar with; namely, a high quality "standard" television picture. This presents difficulty because the standard television picture is most likely much superior to the codec output pictures in many respects. Not only must tests be devised, but a reference against which the evaluation is to be made must also be established. Thus a new evaluation problem exists.

The specific objective of this program is to rank all of the candidate 1.5 Mbps motion codecs as to relative performance by evaluating the quality of the output picture. Based on this requirement, the philosophy proposed is as follows:

- Subjectively evaluate the performance of the codecs, one with respect to the other, to determine which produces the best overall results,

- Generate a performance grade for each of the codecs relative to the best overall performance,

- If desirable, subjectively evaluate the quality of a codec output picture against the input video signal.

The following test outline will recommend and describe the performance of the subjective codec vs codec evaluation. It is presently not anticipated that it will be necessary to perform the third type of tests, namely, evaluating each codec output against its input video signal. Without having performed this test it is difficult to estimate its value since it seems certain that the output picture will be substantially lower in quality than the input picture at least as far as resolution and motion are concerned. However, all of the data necessary to perform this evaluation which may be desirable under special circumstances will be provided.

The testing concepts discussed are presented graphically in Figure 2-1. This figure depicts the conceptual steps in evaluating the performance of motion codecs. The tests consist of two basic parts:

- o Gathering codec performance data, and
- o Ranking the performance of the codecs tested.

Gathering codec performance data consists of preparing a test tape containing the appropriate sequences of video material for subjective testing (such as motion scenes). When the codecs are available, the testing can be carried out on one codec at a time at the manufacturer's facility or other designated location.

The first step consists of feeding the video signal from the test video tape into the codec to be evaluated. The output video signal from the receive side of the codec is recorded on a video tape recorder without performing a grading evaluation at that

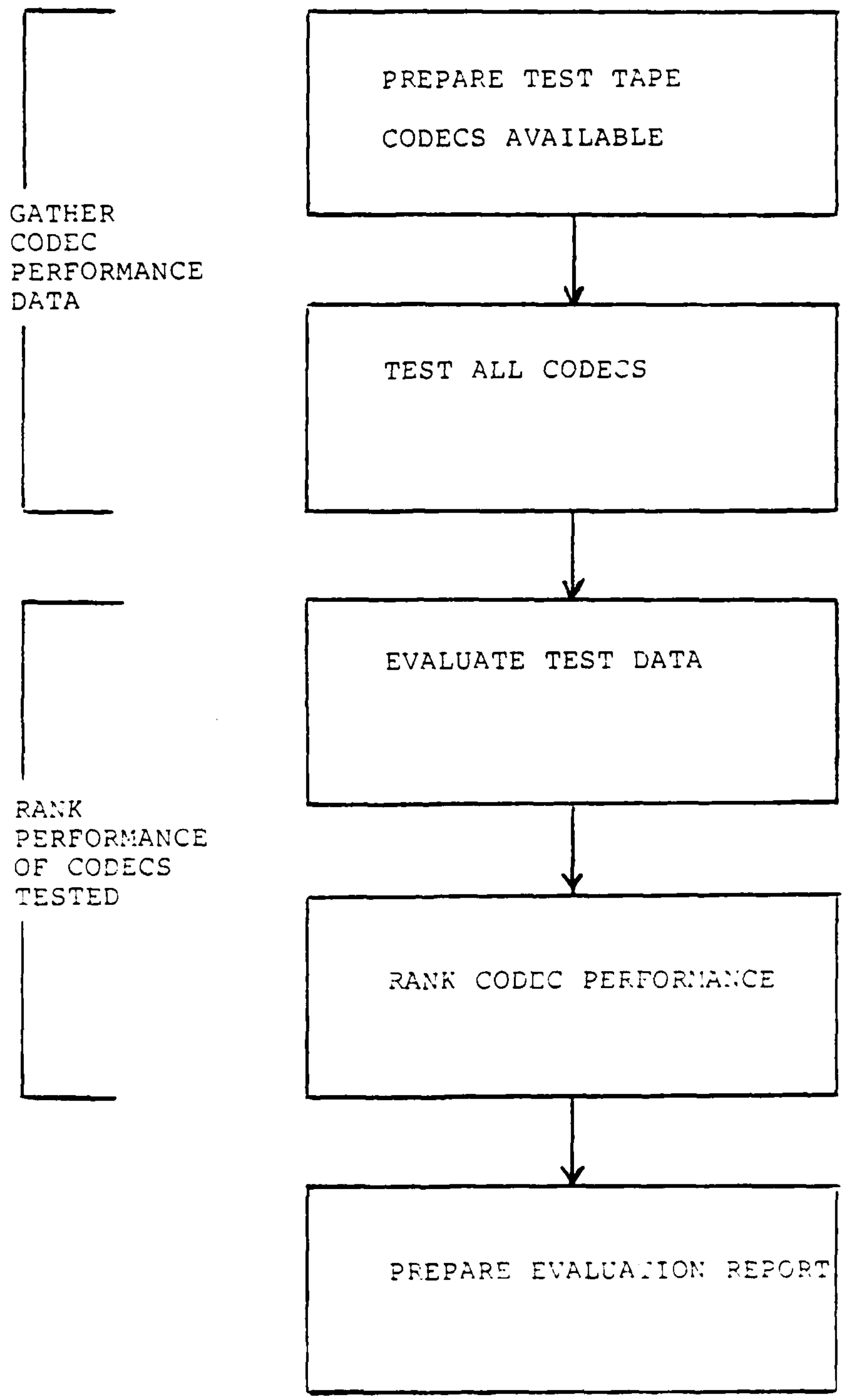


FIGURE 2-1 CODING TESTING CONCEPT

time. Needless to say the video tape recorders must be of high quality so that they will provide an excellent input video signal and will not affect the quality of the recorded output signal.

The second phase of the testing program commences after all of the codec output video tapes of the test pictures have been generated in this way. The second phase will determine a performance grade for each codec in comparison with the other codecs. The test consists of evaluating the performance of each codec as recorded on the video tape against each of the other codecs taken two at a time and determining which performs better. The performance of each codec is ranked as much better, better, slightly better, or the same as, the performance of the codec against which it is being evaluated. Once all of the codecs have been ranked against each other, an overall grade can be developed for each. This process is depicted graphically in more detail in Figure 2-2. The best performing codec is determined as well as the ranking of the other codecs with respect to it.

When this testing procedure has been completed, a report can be generated containing all of the resulting subjective data.

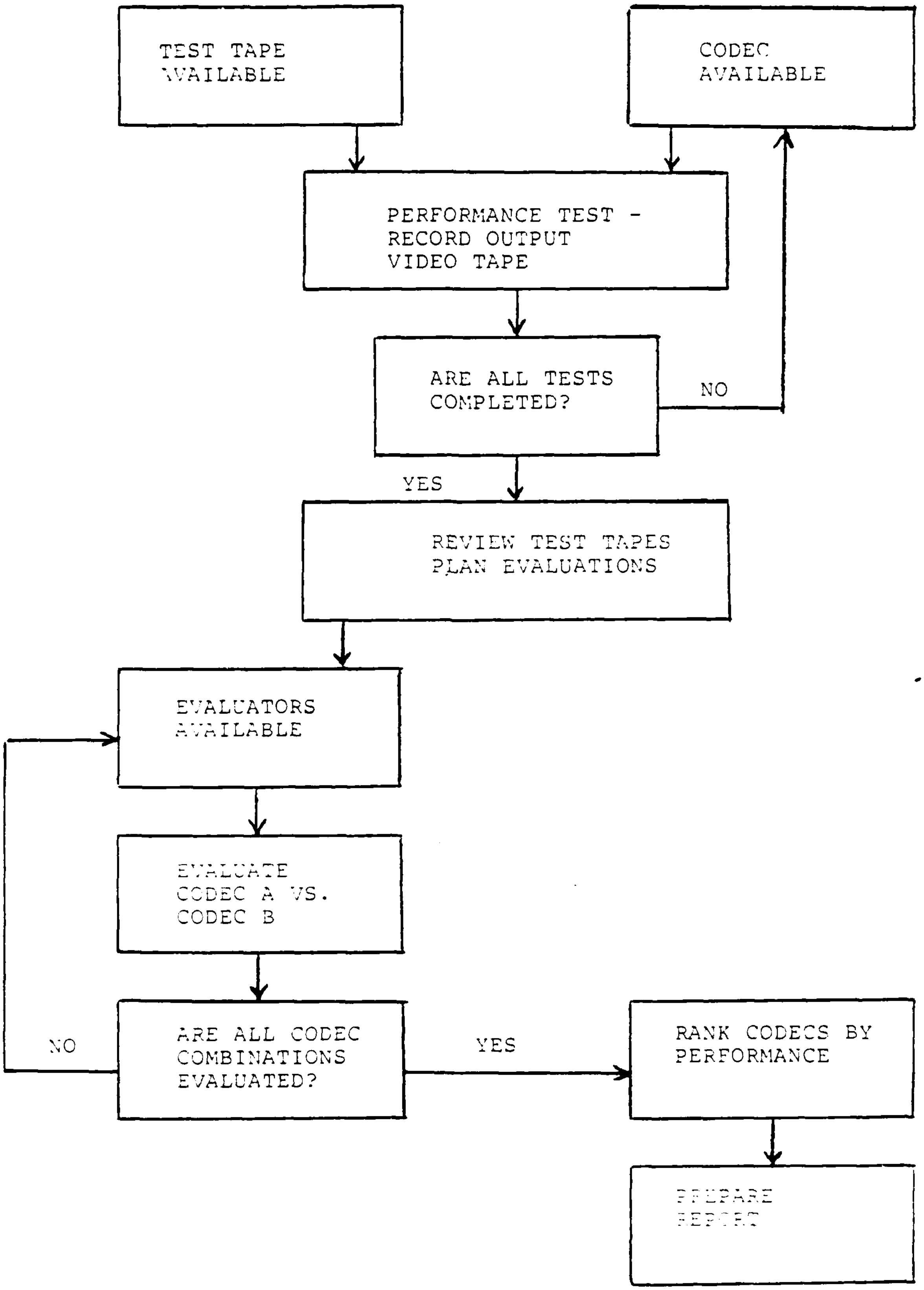


FIGURE 2-2 TESTING PROCEDURE

SECTION 3.0 TESTS TO BE PERFORMED

There are several basic types of tests which are necessary or desirable to completely evaluate the performance of the codecs. They are shown generically in the following tabulation.

TEST	TYPE	REQUIRED
Parameter Measurements		
o Basic Signal Tests	Objective	NO (A,B)
o Video Parameter Tests	Objective	NO (B,C)
o Digital Tests	Objective	NO (B,C)
Performance Evaluation		
o Motion	Subjective	YES
o Quality	Subjective	YES
o Channel Effects	Subjective	YES

A) Accept manufacturer's data for this evaluation

B) Should be included in the acceptance test for each unit (when suitable test procedures have been devised)

C) Desirable tests to gather data to correlate with the subjective performance grade. Test signals will be provided, passed through the codecs and recorded to permit this evaluation in the future.

3.1 PARAMETER MEASUREMENTS

Table 3-1 lists in detail all of the parameters included in the above generic categories and assigns to each a level of test priority. They include the various tests which are normally

TABLE 3-1, TEST PRIORITY

TITLE OF TEST BASIC SYSTEM TESTS	TEST LEVEL
1.0 VIDEO AMPLITUDE	
1.1 OVERALL AMPLITUDE (INSERTION GAIN)	A,B
1.2 SYNC AMPLITUDE	A,B
1.3 BURST AMPLITUDE	A,B
1.4 SETUP	A,B
2.0 SYNC TIMING MEASUREMENTS	
2.1 SYNC FORMAT	A,B
2.2 VERTICAL BLANKING	A,B
2.3 EQUALIZING PULSE WIDTH	A,B
2.4 VERTICAL SYNC PULSE	A,B
2.5 VERTICAL SERRATION WIDTH	A,B
2.6 HORIZONTAL BLANKING	A,B
2.7 FRONT PORCH WIDTH	A,B
2.8 BURST	A,B
2.9 BREEZEWAY	A,B
2.10 RISE AND FALL OF H. SYNC	A,B
2.11 SUBCARRIER FREQUENCY	A,B
3.0 OTHER BASIC TESTS	
3.1 INPUT IMPEDANCE (RETURN LOSS)	A,B
3.2 LOAD IMPEDANCE	A,B
3.3 OUTPUT IMPEDANCE (RETURN LOSS)	A,B
3.4 POLARITY OF PICTURE SIGNAL	A,B
3.5 NON-USEFULL D-C COMPONENT	A,B

TITLE OF TEST DIGITAL/CHANNEL TESTS	TEST LEVEL
1.0 DIGITIZATION TEST	
1.1 FILTER PARAMETERS	B,C
1.2 SAMPLING RATE EFFECTS	
1.2.1. LUMINANCE	B,C
1.2.2 CHROMINANCE	B,C
1.3 SAMPLING PRECISION EFFECTS	
1.3.1 LUMINANCE	B,C
1.3.2 CHROMINANCE	B,C
1.4 LINEAR DISTORTION	3
1.5 NON-LINEAR DISTORTION	4
2.0 SIGNAL PROCESSING	
2.1 COMPRESSION ALGORITHM	
2.1.1 LUMINANCE	S
2.1.2 CHROMINANCE	S
2.2 MOTION CAPABILITY	S,S
2.3 SYNCHRONIZATION SCHEME	A,B
3.0 CHANNEL EFFECTS	
3.1 BIT ERRORS	S
3.2 ERROR DISTRIBUTION	S

TITLE OF TEST VIDEO TESTS	TEST LEVEL
1.0 LINEAR-DISTORTION	
1.1 AMPLITUDE VS. FREQUENCY CHARACTERISTICS	B,C,1
1.2 LINEAR CHROMINANCE DISTORTION	
1.2.1 CHROMINANCE-TO-LUMINANCE GAIN INEQUALITY	B,C
1.2.2 CHROMINANCE-TO-LUMINANCE DELAY INEQUALITY	B,C
1.3 ENVELOPE DELAY VS FREQ. CHARACTERISTICS	2
1.4 FIELD TIME WAVEFORM DISTORTION	A,B
1.5 LINE TIME WAVEFORM DISTORTION	A,B
1.6 SHORT TIME WAVEFORM DISTORTION	A,B
1.7 LONG TIME WAVEFORM DISTORTION (BOUNCE)	A,B
1.8 INSERTION GAIN AND GAIN VARIATION	A,B

TITLE OF TEST VIDEO TESTS	TEST LEVEL
2.0 NONLINEAR DISTORTION	
2.1 LUMINANCE NONLINEARITY	B,C
2.2 LUMINANCE-TO-CHROMINANCE INTERMODULATION	
2.2.1 DIFFERENTIAL GAIN	B,C
2.2.2 DIFFERENTIAL PHASE	B,C
2.3 CHROMINANCE-TO-LUMINANCE INTERMODULATION	B,C
2.4 CHROMINANCE NONLINEARITY	
2.4.1 CHROMINANCE NONLINEAR GAIN	B,C
2.4.2 CHROMINANCE NONLINEAR PHASE	B,C
2.5 DYNAMIC GAIN	
2.5.1 DYNAMIC GAIN OF PICTURE SIGNAL	B,C
2.5.2 DYNAMIC GAIN OF SYNCH-RONIZING SIGNAL	A,B
2.6 TRANSIENT SYNCHRONIZING SIGNAL NONLINEARITY	A,B

TITLE OF TEST VIDEO TESTS	TEST LEVEL
3.0 INTERFERENCE	
3.1 SIGNAL-TO-NOISE RATIO (10 KHZ - 5.0 KHZ)	A,B
3.2 SIGNAL-TO-LOW FREQUENCY NOISE RATIO (0 - 10 KHZ)	A,B
3.3 SIGNAL-TO-PERIODIC NOISE RATIO (300 HZ - 4.2 MHZ)	A,B
3.4 SIGNAL-TO-IMPULSE NOISE RATIO	A,B
4.0 CONTINUITY OF VIDEO SERVICE	A

A; ACCEPT MANUFACTURER'S DATA FOR THIS EVALUATION.
B; ACCEPTANCE TEST FOR EACH UNIT.
C; DESIRABLE TEST TO GATHER OBJECTIVE DATA TO CORRELATE WITH SUBJECTIVE PERFORMANCE GRADE.
TEST SIGNALS WILL BE PROVIDED, PASSED THROUGH THE CODEC,AND RECORDED, TO PERMIT FUTURE EVALUATION.
S; SUBJECTIVE TEST
1; SEE ALSO DIGITAL/CHANNEL TESTS 1.1, 1.2.1, AND 1.2.2.
2; LINEAR DISTORTION TEST 1.2.2 CHROMINANCE TO LUMINANCE DELAY INEQUALITY WILL SUFFICE. OFFICIAL REQUIREMENT STILL TO BE DEFINED.
3; SEE VIDEO TESTS, 1.0 LINEAR DISTORTION.
4; SEE VIDEO TESTS, 2.0 NON-LINEAR DISTORTION.
5; OBJECTIVE TEST METHODOLOGY STILL TO BE DEVELOPED.

performed on video systems. It is not considered a requirement nor is it deemed essential to the performance of this program that these parameter tests be conducted at this time. However, test signals will be provided and transmitted through each codec pair. The output video of these test signals will be recorded for future evaluation. Correlation of this data with the subjective tests will be an aid in developing objective tests specifically for teleconferencing type codecs. The test signals provided will accomodate all C level and most B level tests. The C level tests include the following:

- o Luminance amplitude vs frequency response
- o Chrominance to luminance gain inequality
- o Chrominance to luminance delay inequality
- o Luminance non-linearity
- o Differential gain
- o Differential phase
- o Chrominance to luminance intermodulation
- o Chrominance non-linear gain
- o Chrominance non-linear phase
- o Dynamic gain of picture signal
- o Luminance filter parameters
- o Luminance sampling rate
- o Chrominance sampling rate
- o Luminance sampling precision
- o Chrominance sampling precision

The test signals used will include those presently considered most informative in video channel testing. Tentatively, test

signals listed below will be included. The white/black window signal and the yellow/blue signal are included to permit the objective measurement of codec motion capability in the future. For a description of this concept refer to Appendix A.

- o Composite test signal, containing modulated 12.5T (T is 125 nanoseconds) sine squared pulse and 18 microsecond bar with sync and blanking
- o Video sweep with sync and blanking (or multiburst)
- o Five or ten step staircase signal, both unmodulated and modulated, with sync and blanking with APL (average picture level) of 10, 50, and 90%.
- o Modulated and unmodulated ramp with sync and blanking
- o Three level chrominance signal with sync and blanking
- o White/black window
- o Yellow/blue signal

More details of the test signals and their purpose are listed in Table 3-2.

3.2 PERFORMANCE EVALUATION

3.2.1 Motion

The remaining tests are those considered most important to the evaluation of the codecs. They will all be implemented subjectively, comparing the performance of one codec against another in all combinations. The motion performance will be evaluated using scenes containing zooming and panning, and others in which the scene itself contains the motion such as moving

TABLE 3-2

TEST SIGNALS

Description	Purpose
Composite Test Signal - consisting of modulated 12.5T sine square pulse, 2T pulse, vertical white bar	Chrominance-to luminance gain and delay inequality, line time and short time waveform distortion
Video Sweep (with markers)	Amplitude vs. frequency response, filter parameters, luminance and chrominance sampling rates.
5 Step staircase, APL=50% 5 Step staircase, APL=90% 5 Step staircase, APL=10%	Luminance nonlinearity Dynamic gain of picture signal and sync signal
5 Step modulated staircase, APL= 50% 5 Step modulated staircase, APL= 90% 5 Step modulated staircase, APL= 10%	Differential gain Differential phase
Ramp	Luminance sampling precision
Modulated Ramp	Chrominance sampling precision
3 Level Chroma	Chrominance-to-luminance intermodulation Chrominance nonlinear gain and phase
Switch between white window & black field, 3 times 10 seconds each.	Objective motion measurement
Switch between yellow & blue field, 3 times 10 seconds each.	Objective or subjective motion response evaluation.

subjects. The motion capability will probably manifest itself in one or more ways such as the following:

- o Time required to update a single, rapid scene change.
- o Distortion of the presentation within the change area in less violently changing scenes.
- o Distortion and spurious patterns (artifacts) within the static area of changing scenes.
- o Reduced overall quality during motion scenes.

Appendix A contains a description of a concept for the objective measurement of codec motion capability. The signals required for this evaluation have been included on the test tape and will be passed through the codecs to permit future evaluation of the codec motion capability as well as the measurement concept itself.

3.2.2 Still Imagery Quality

The quality performance of the codecs will be evaluated by using scenes specifically designed to contain features which stress certain functions of the codecs. For example, scenes containing text of various sizes, fonts, and luminance/chrominance combinations stress resolution and transfer characteristics of the codecs. Patterns in the scene background and in the subjects' apparel providing gradually varying colors and gray levels, sharp transitions, and high contrast patterns all yield information which can be subjectively and comparatively evaluated. These selected scenes will aid in evaluating many of the technical parameters listed in Table 3-1. Samples of these parameters

include the following:

- o Aliasing.
- o Sampling rate and precision for both luminance and chrominance.
- o Linear and non-linear distortion
- o Compression algorithm.

3.2.3 Channel Effects

Channel effects will be evaluated by contaminating the digital transmission of selected scenes between the transmit codec and the receive codec with bit errors at accurately known rates. The results of this process will be an integral part of the recording of the output video signal of the various codecs being evaluated. The evaluation will again be on a comparative basis with other codecs. The factors which present themselves in the output picture due to the channel errors include the following examples.

- o Deterioration of picture quality.
- o Artifacts in the contaminated video frames.
- o Time required to correct these artifacts, etc.
- o Loss of sync.

This further evaluates the codec compression algorithm and the error correction capability.

SECTION 4 PROCEDURE TO GATHER DATA

The data gathering phase of the test program consists of passing specially designed video signals through the codec pair (transmitter and receiver) and recording the output which consists of the picture sequences recorded on video tape for subjective evaluation and the video test signals for future objective evaluation if and when desired.

Figure 4-1 is a block diagram showing the test implementation. The signal source, a video tape recorder, is connected to the codec transmitter to be tested. The video signal from the video tape recorder is monitored for quality and level on a television picture monitor and on a television waveform monitor/vectorscope. The digital signal from the codec transmitter is directly connected to the codec receiver in a normal configuration. Only for the test which determines the effect of channel errors on codec performance, this connection is made through a bit error inserter. The receive codec is connected to a television waveform monitor/vectorscope, a television picture monitor, and a high quality video tape recorder. Note that level, impedance and termination criteria must be carefully observed.

The video signal is then passed through the codec pair and the output of the receive codec is recorded on the second video tape recorder. The video tape thus generated containing the test video signals as they are reconstructed by the receive codec will be evaluated to determine the performance of the codec subjectively and possibly later objectively in conventional analog terms.

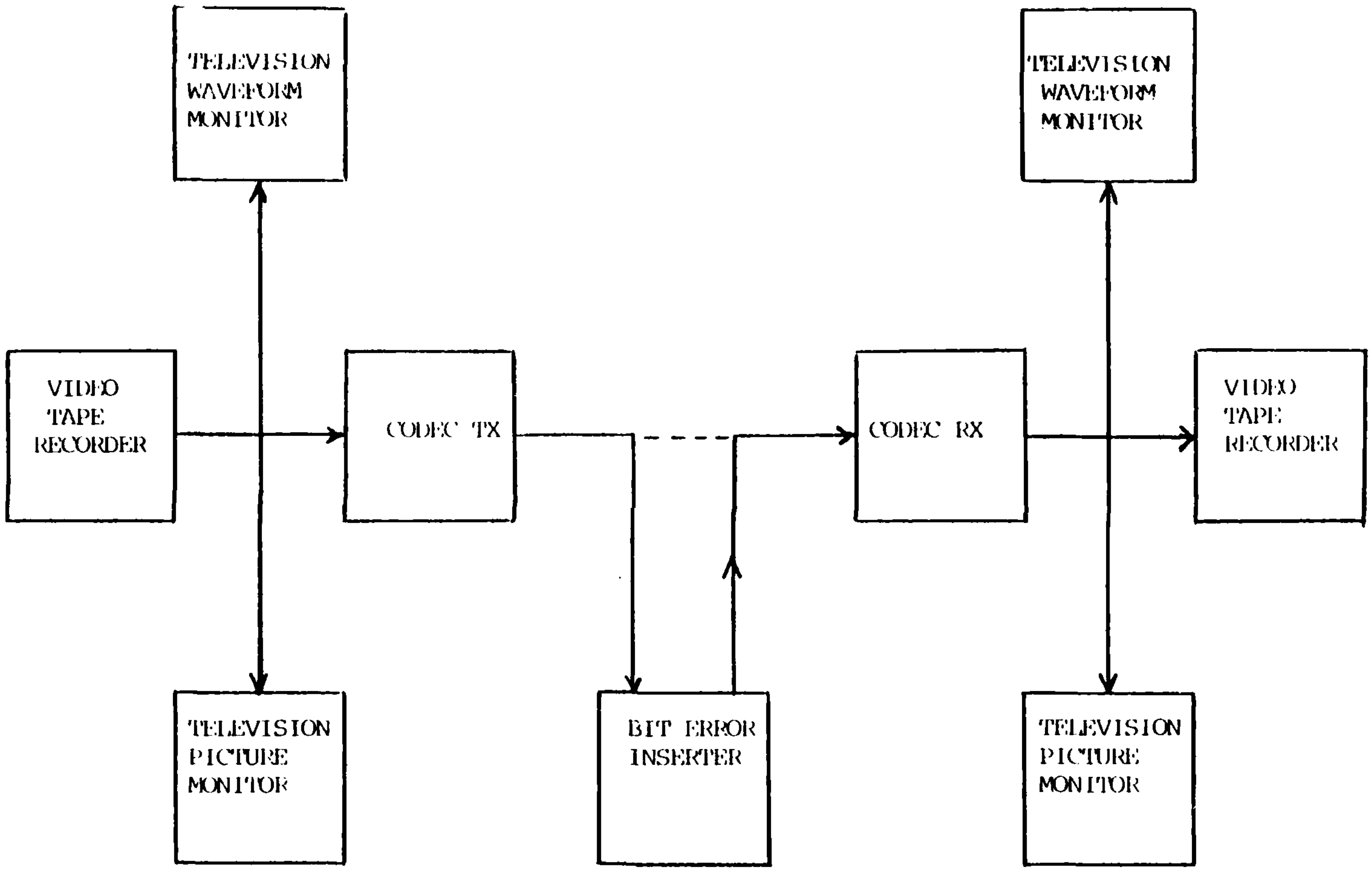


FIGURE 4-1 TEST IMPLEMENTATION AT THE MANUFACTURER"S FACILITY

SECTION 5 EVALUATION PROCEDURE

5.1 Evaluation Philosophy

The philosophy of evaluation is based on the fact that the category of video transmission equipments being evaluated are fairly new. Universally accepted objective tests to determine their performance have not yet been developed. Therefore, the performance must be evaluated subjectively.

The quality of the output picture is, at least in some aspects, substantially different from that of the "Broadcast quality" picture. Broadcast quality transmission systems can be precisely defined by means of a set of specifications which can be very accurately measured objectively. This is possible only because of a considerable amount of effort in the infancy of television broadcasting which correlated the subjective evaluation of pictures transmitted through a channel (analog in this case) with the results of measurements of the physical parameters of that channel. As a result, the subjective quality of the received picture can be predicted based on an objective evaluation of the transmission channel.

The modern 1.544 Mbps (or lower) codec has not yet been through this growth phase. It must be remembered that the codec pair, transmitter and receiver, have a combined transfer function and thus constitute a transmission system. When used with a digital channel of adequate performance (low error rate), the channel essentially drops out of consideration because it does not deteriorate the signal. The codec alone determines performance.

If the digital channel is not error free, the amount of degradation of the output picture due to errors in the received digital signal is again determined by the codec, in particular by the compression algorithm and error correction scheme employed. Therefore, the codec pair can be treated and tested as a transmission system.

Several methods of subjectively evaluating the codec performance are possible as described below.

5.1.1 Absolute Rating

The absolute method of evaluating the output picture requires the evaluator to give a rating to the quality of the picture as an entity in itself without relating it to any other picture. This is akin to rating the frequency of a tone or the color of an object on an absolute basis without a reference. Many musicians have perfect pitch recognition but no visual equivalent to this capability is known. Human beings tend to perform almost all evaluations on a comparative basis; "This is better than that, This is much poorer than that", etc. It is very difficult to give absolute ratings with any degree of consistency. Evaluators almost always make a decision in reference to some standard whether intentionally or not. It is anticipated that, since the evaluators will not be familiar with this type of equipment, they will relate the output picture to the television pictures they are most familiar with; namely, high quality broadcast television pictures. It is anticipated that an attempt at absolute evaluation of the output picture may produce inconsistent and

highly diverse performance grades which will not be correlatable to produce the desired final codec rating. To avoid this possible pitfall, absolute evaluation is not recommended for quality assessment.

5.1.2 Comparison With Input Data

Comparing the output picture with the input picture is a viable candidate for the evaluation technique. It is a valid approach from the standpoint of producing meaningful and, most probably, consistent results. CCIR REC 500-2 lists the rating scales shown in Figure 5-1.

Two factors should be considered, however. First, the quality of the output signal will most likely always be substantially poorer that the quality of the input signal. Therefore, the comparison scale is quite restrictive because only half of its grades relate to lower quality while the rest relate to equal or higher quality. A grading scale such as the impairment scale (Figure 5-1) would be more useful since it has 5 impairment grades. However it is quite possible that to the non-professional evaluator the output picture will always contain impairments in the "annoying" or "very annoying" categories. If evaluation were on a resolution basis alone and the codec maximum output resolution is "256 x 256" picture elements as compared to the input resolution of "512 x 512" it is conceivable that a grade of no better than "annoying" would always result. Therefore the impairment grading scale could be extremely restrictive to the point that only two useful grades might apply. While a very

IMPAIRMENT SCALE

THE IMPAIRMENT OR QUALITY OF THE LEFT PICTURE AS COMPARED TO THE RIGHT PICTURE

GRADE	IMPAIRMENT	QUALITY
5	IMPERCEPTIBLE	EXCELLENT
4	PERCEPTIBLE, BUT NOT ANNOYING	GOOD
3	SLIGHTLY ANNOYING	FAIR
2	ANNOYING	POOR
1	VERY ANNOYING	BAD

COMPARISON SCALE

THE QUALITY OF THE LEFT PICTURE AS COMPARED TO THE RIGHT PICTURE

GRADE	QUALITY
+3	MUCH BETTER
+2	BETTER
+1	SLIGHTLY BETTER
0	THE SAME
-1	SLIGHTLY WORSE
-2	WORSE
-3	MUCH WORSE

FIGURE 5-1 SUBJECTIVE GRADING SCALES

important factor would clearly be defined, namely that the quality of the output with respect to the input is always substantially poorer, it may be very difficult to grade the performance of one codec with respect to the other. It is, however, precisely this relative evaluation which is most useful and the specific goal of the evaluation process.

Therefore, comparison of the output video signal with the input video signal is not recommended as the primary evaluation method but it may be useful in special cases, for instance to resolve an ambiguity. Furthermore, if the number of codecs to be ranked is large, the number of required paired comparison tests becomes unwieldy. In this case it is recommended to first compare each codec output with the input and then hold a "runoff" by paired comparsions between the 2 to 4 codecs having shown the least impairments.

5.1.3 Comparison Among Codecs

The technique of comparison between codecs appears to provide the best method of grading the codecs, each with respect to all of the others. In this approach each codec is graded against each of the other codecs. Figure 5-2 shows the concept for a total of 5 codecs. Evaluating the 5 codecs against each other requires a total of 10 evaluation tests. This is double the number of tests required by either the absolute approach or the comparison with the input picture approach. However, since the pictures being evaluated in this test are both the output of codecs which cause degradation, their quality is likely to be fairly similar. The

	LEFT MONITOR	RIGHT MONITOR
TEST	CODEC NO.	CODEC NO.
1	1	2
2	1	3
3	1	4
4	1	5
5	2	3
6	2	4
7	2	5
8	3	4
9	3	5
10	4	5

TOTAL NUMBER OF TESTS REQUIRED = (N (N-1))/2

FIGURE 5-2 COMPARISON TESTS

evaluators will be able to use a much greater part of the grading scale than in the previous approach which compared the quality of the output signal with that of the input signal. It is felt that this will better achieve the goal of ranking the codecs for quality performance.

The grading scale shown on the codec evaluation form, Figure 5-3, is recommended. Since either of the codecs being evaluated can perform better than the other in any specific parameter, a scale which can rate either picture better than the other is necessary. The comparison scale of Figure 5-3 has this feature.

5.2 Evaluation Procedure

The basic concept of the evaluation procedure is very simple. It is shown in Figure 5-4. Two specifically prepared video tapes, each containing the same output pictures from different codecs in the same sequence, are displayed each on separate television monitors. The evaluators compare the quality of the two pictures and grade them on a comparative basis. The figure is deceivingly simple but each parameter of the test must be very carefully controlled to assure valid evaluation results. Since it is known that adjusting side-by-side color monitors for exactly equal performance is very difficult, a monitor reversal switch is provided so that the effect of any small difference between monitors can be eliminated.

The characteristics of the viewing area must be controlled. Table 5-1 lists the more inportant recommended viewing conditions. The tabulation is an excerpt from CCIR REC 500-2.

CODEC EVALUATION FORM

EVALUATOR: ______________________________ TEST NO: ______________________________

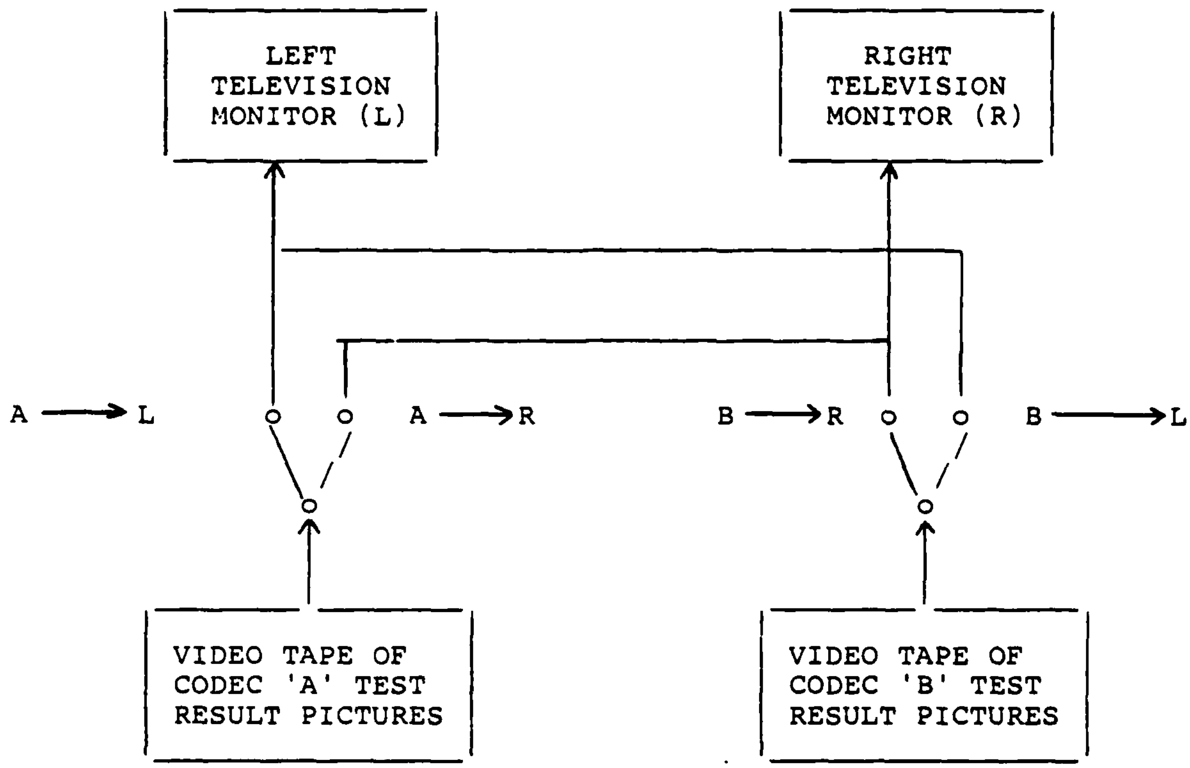
DATE: ______________________________ PAGE: ______________________________

+3; MUCH BETTER THAN	+2: BETTER THAN	+1; SLIGHTLY BETTER THAN	0; SAME AS

TEST SEQ.	LEFT PICTURE BETTER			SAME AS	RIGHT PICTURE BETTER		
1	+3	+2	+1	0	+1	+2	+3
2	+3	+2	+1	0	+1	+2	+3
3	+3	+2	+1	0	+1	+2	+3
4	+3	+2	+1	0	+1	+2	+3
5	+3	+2	+1	0	+1	+2	+3
6	+3	+2	+1	0	+1	+2	+3
7	+3	+2	+1	0	+1	+2	+3
8	+3	+2	+1	0	+1	+2	+3
9	+3	+2	+1	0	+1	+2	+3
10	+3	+2	+1	0	+1	+2	+3
11	+3	+2	+1	0	+1	+2	+3
12	+3	+2	+1	0	+1	+2	+3
13	+3	+2	+1	0	+1	+2	+3
14	+3	+2	+1	0	+1	+2	+3
15	+3	+2	+1	0	+1	+2	+3

TEST SEQ.	LEFT PICTURE BETTER			SAME AS	RIGHT PICTURE BETTER		
16	+3	+2	+1	0	+1	+2	+3
17	+3	+2	+1	0	+1	+2	+3
18	+3	+2	+1	0	+1	+2	+3
19	+3	+2	+1	0	+1	+2	+3
20	+3	+2	+1	0	+1	+2	+3
21	+3	+2	+1	0	+1	+2	+3
22	+3	+2	+1	0	+1	+2	+3
23	+3	+2	+1	0	+1	+2	+3
24	+3	+2	+1	0	+1	+2	+3
25	+3	+2	+1	0	+1	+2	+3
26	+3	+2	+1	0	+1	+2	+3
27	+3	+2	+1	0	+1	+2	+3
28	+3	+2	+1	0	+1	+2	+3
29	+3	+2	+1	0	+1	+2	+3
30	+3	+2	+1	0	+1	+2	+3

FIGURE 5-3, CODEC EVALUATION FORM



EQUIPMENT SETUP FOR EVALUATION OF CODEC PERFORMANCE, IN PAIRS. BY COMPARISON OF OUTPUT PICTURES.

FIGURE 5-4 EVALUATION SETUP

TABLE 5-1 RECOMMENDED SUBJECTIVE VIEWING CONDITIONS

PARAMETER	RECOMMENDATION (CCIR 500)
RATIO OF VIEWING DISTANCE TO PICTURE HEIGHT	4 TO 6
PEAK SCREEN LUMINANCE	70 CD/SQ. M
RATIO OF INACTIVE SCREEN TO PEAK LIMINANCE	<0.02
RATIO OF BACKGROUND LUMINANCE TO PEAK SCREEN LUMINANCE	0.15
AMBIENT ILLUMINANCE	LOW
CHROMATICITY OF SURROUND	D65

The specific format of the video tape is shown in Figure 5-5. Each video tape consists of the same sequence of output pictures, each recorded through a different codec. The pictures are interspersed with neutral identification frames. The following describes the comparison procedure. Picture 'N' from codec 'A' is to be compared with picture 'N' from codec 'B'. The picture from codec 'A' is presented on the left hand monitor: The picture from codec 'B' is presented on the right hand monitor. They are displayed for a period of time to permit adequate comparison, not less than about 15 seconds, followed by a 10 second display of a neutral field containing the caption "Score Sequence N". This will permit the evaluators adequate time to record the grade they have assigned to the better performing codec. The process is repeated for all pictures from 1 to N as shown in Figure 5-6.

The switch which is provided to reverse the physical order of the displays permits showing a duplicate of a previously presented picture on the opposite monitor; that is, the picture previously displayed on the right hand monitor now appears on the left and vice versa. This is shown for two selected sequences in Figure 5-6. These sequences are also graded by the evaluators. Correlating the results of the reversed display with the original will provide an additional degree of confidence in the tests. Display reversal for half of the test tape is also possible. The total test sequence should be limited to about 30 minutes.

Figure 5-3 shows the suggested codec evaluation form on which the evaluators record the grade which they assign to each sequence. It contains headers and spaces for all of the pertinent

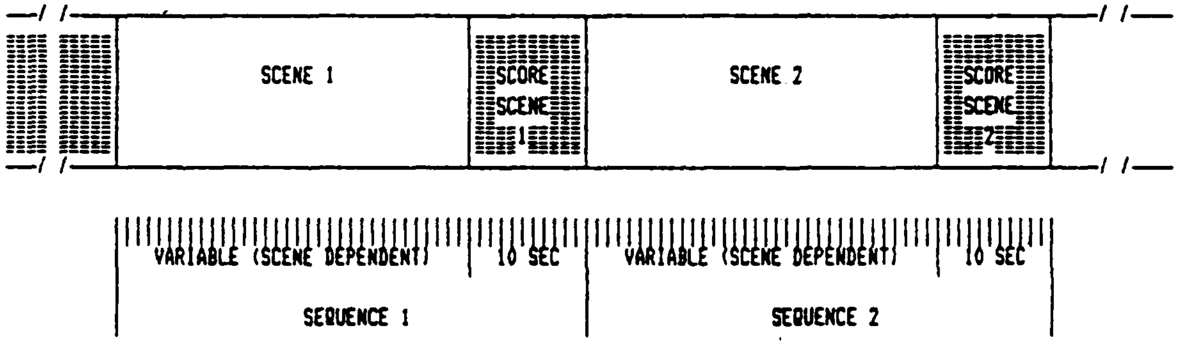


FIGURE 5-5, STIMULUS PRESENTATION

SEQUENCE	PICTURE	CODEC	MONITOR
1	1 1	A B	LEFT RIGHT
2	2 2	A B	LEFT RIGHT
(N-1)	(N-1) (N-1)	A B	LEFT RIGHT
(N)	(N) (N)	A B	LEFT RIGHT
CHECK 1	SELECT 1 SELECT 1	A B	RIGHT LEFT
CHECK 2	SELECT 2 SELECT 2	A B	RIGHT LEFT

THIS SEQUENCE PERMITS THE COMPARATIVE EVALUATION (GRADING) OF THE RELATIVE PERFORMANCE OF TWO CODECS. EACH PICTURE ON THE TEST TAPE (1 TO N) HAS BEEN TRANSMITTED THROUGH EACH CODEC, A AND B, THE OUTPUT PICTURE FROM CODEC A IS PRESENTED ON THE LEFT MONITOR: THE OUTPUT PICTURE FROM CODEC B IS PRESENTED ON THE RIGHT MONITOR. THE EVALUATORS SELECT THE HIGHER QUALITY PICTURE AND GRADE IT ON A COMPARATIVE BASIS. SEVERAL CHECK SEQUENCES ARE PROVIDED (TWO ARE SHOWN IN THE FIGURE ABOVE). IN THESE SEQUENCES PICTURES ALREADY PRESENTED ARE SHOWN BUT WITH THE CODEC/MONITOR RELATIONSHIP REVERSED. THIS WILL PROVIDE AN ADDED DEGREE OF CONFIDENCE IN THE TEST RESULTS. THESE CHECK SEQUENCES MAY BE INTERSPERSED WITH OTHERS AND MONITOR REVERSAL FOR HALF OF THE TAPE MAY BE USED.

FIGURE 5-6 EVALUATION SEQUENCES

data to define the test run and the evaluator. The grading scale is printed on this page to serve as a reference should the evaluator need it. The recording procedure is to place a mark in the box containing the appropriate grade for each sequence. This format assures consistent recording of grades with an absolute minimum of distraction for the evaluator.

5.3 Evaluation Computation

The preceding sections described the method of generating data from which a quantitative evaluation of codec performance can be ascertained. The following is a description of the calculations which produce a single quantitative grade for a codec's performance as compared to the performance of similar codecs.

The concept of comparing codecs A and B is shown in Figure 5-7. The major matrix in this figure is a planar matrix which lists the sequences evaluated along the ordinate and the evaluators along the abscissa. This matrix is a graphic presentation of the relationship of the calculations to the evaluation form of Figure 5-3.

The first calculation determines a mean and a standard deviation for each sequence as indicated by the arrows. The mean indicates the comparative performance of the codecs for each sequence. Since the reaction of each codec to different sequences is completely variable, the mean values may cover the whole range from +3 to -3 and be entirely valid. However, a high standard deviation indicates wide disagreement between evaluators. Should

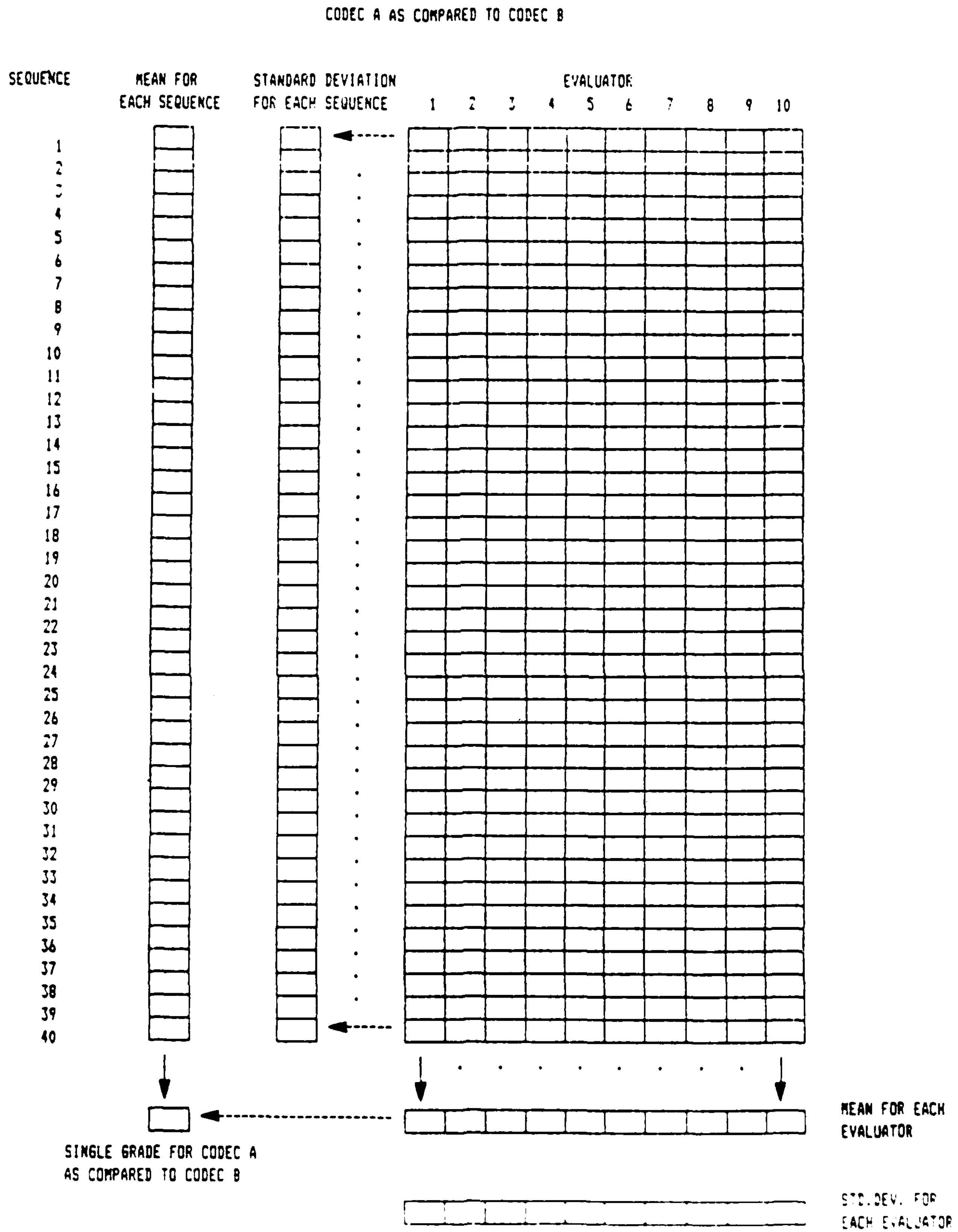


FIGURE 5-7, CODEC A AS COMPARED TO CODEC B

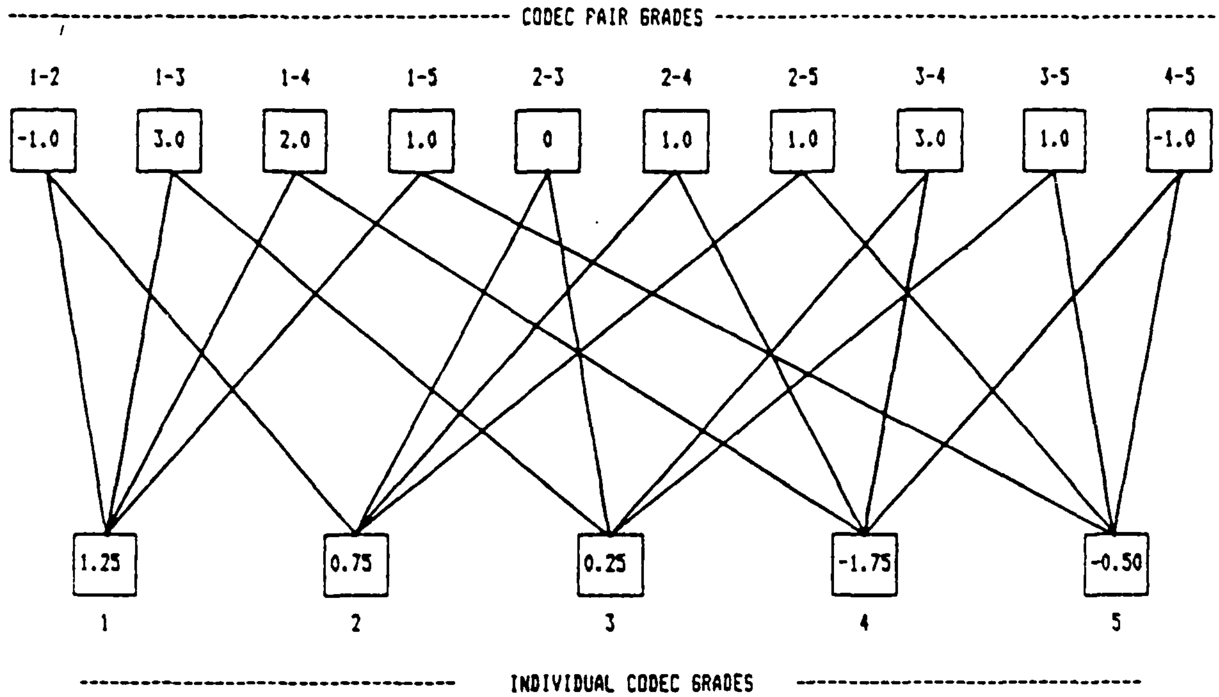
this occur for a specific sequence in several codec comparisons, it shows that scoring of this sequence is unduly difficult and it may be advisable to exclude it from the evaluation.

The second calculation is to determine the mean and the standard deviation for each of the evaluators. The rationale is similar to before: specifically, a mean far out of line with the means of the other evaluators provides some concern as to the meaningfulness of the scoring by that particular evaluator, particularly if it recurs in several codec comparisons. The standard deviation, however, is determined by the range of scores used by each evaluator. Therefore, a high standard deviation mainly indicates a very outspoken and determined evaluator and by itself is no reason to question the validity of the result unless a close scrutiny of the scores indicates that this evaluator is just entering arbitrary nonsense numbers.

It remains then to produce a single grade for that specific codec comparison; eg., codec A compared to codec B. That single grade is determined as the mean of the means of the individual evaluators or of all the sequences employed. Both calculations cover all evaluators and test sequences and must yield the same grade. If this grade is positive, codec A has performed better than codec B: If this grade is negative, codec B has performed better than codec A. In this manner a single grade can be developed for each codec comparison.

The next set of calculations will rank the codecs. This is depicted graphically in Figure 5-8. To do this, a single grade

EVALUATION CALCULATION CONCEPT



RANK	CODEC	GRADE
1	1	1.25
2	2	0.75
3	3	0.25
4	5	-0.50
5	4	-1.75

FIGURE 5-8, EVALUATION CALCULATION CONCEPT

must be developed to indicate the performance of each codec as compared to all other codecs. If, for example, there are 5 codecs, a single grade must be developed for the performance of codec A as compared to codec B, C, D, and E. In the previous paragraph the method of determining a grade for the evaluation of codec A as compared to codec B was presented. By extension, this same technique is employed to determine a single grade for the performance of codec A as compared to codec C, A to D, and A to E. It follows then that the single grade for the performance of codec A as compared to all other codecs is the mean of these individual grades. This same procedure applies to determining a single performance grade for codecs B, C, D, and E. Note that the grade of codec B as compared to codec A is the negative of the grade for codec A as compared to codec B. Figure 5-8 shows the 10 scores of individual codec pairs on top. The single grades for each codec are shown in the 5 boxes below, with the connecting lines indicating how each of the single grades was derived as the mean of 4 individual scores.

The ranking procedure is now simply a case of ranking the overall performance grades of the individual codecs. However, retaining the individual grades will assist in determining how much better any one codec performed as compared to the total number of codecs.

The arbitrary scores used herein happened to produce an ambiguity which, though highly unlikely, may occur during an actual codec evaluation. The final ranking shows codec 1 first and codec 2 second but in the paired comparison codec 2 was scored

better than codec 1. Such a case would require a detailed analysis and possibly additional scoring. Separate evaluations by several expert evaluators may be in order, or comparison of the codec output signals with the input signal, using the impairment scale, may yield sufficient added data to resolve the ambiguity.

Figures 5-9 and 5-10 are the printouts produced from a very straightforward computer program which implements the philosophy described above. Figure 5-9 is the data entry form. It shows the data from the individual evaluators generated in the evaluation of the performance of codec A as compared to codec B entered into a single file for further computation. Ten evaluators and 40 sequences were assumed. The mean and standard deviation for each sequence are printed on the left in columns 2 and 3. The mean and standard deviation for each evaluator are printed in clearly marked horizontal rows. Figure 5-10 shows the basic calculation involved; namely, the conversion from two columns of entries per evaluator (right and left) to one column in which the entries which were in the right column are entered as negative numbers. This is a totally consistent approach because Codec A (on the left) is being compared with Codec B (on the right). Therefore, a positive numerical evaluation in any column is equivalent to a negative evaluation in the other column. This conversion is made in Figure 5-10 and is included here for completeness.

Specific items in these figures have been circled as items of interest. The correlation of the sample data entered in these figures to data anticipated in the actual evaluations is, at this point, conjecture, and not important to demonstrate the validity

CODEC EVALUATION

LEFT CODEC : A   MFG. BY: SUPERTECH   GRADE : 0.928
RIGHT CODEC : B   MFG. BY: DIGIVIDEO   GRADE : -0.928

	EVALUATORS 1	2	3	4	5	6	7	8	9	10
MEAN :	1.525	1.375	1.525	1.375	1.525	1.400	1.550	0.500	0.000	-1.500
STD DEV:	1.154	1.334	1.154	1.334	1.154	1.336	1.154	0.877	3.038	0.506

SEQ.	MEAN	STD. DEV.	1 L	1 R	2 L	2 R	3 L	3 R	4 L	4 R	5 L	5 R	6 L	6 R	7 L	7 R	8 L	8 R	9 L	9 R	10 L	10 R
1	2.20	1.32	3	0	3	0	3	0	3	0	3	0	2	0	2	0	1	0	3	0	0	1
2	2.20	1.32	3	0	3	0	3	0	3	0	3	0	2	0	2	0	1	0	3	0	0	1
3	2.10	1.29	3	0	3	0	3	0	3	0	2	0	2	0	2	0	1	0	3	0	0	1
4	2.10	1.29	3	0	3	0	3	0	3	0	2	0	2	0	2	0	1	0	3	0	0	1
5	2.00	1.25	3	0	3	0	3	0	2	0	2	0	2	0	2	0	1	0	3	0	0	1
6	2.00	1.25	3	0	3	0	3	0	2	0	2	0	2	0	2	0	1	0	3	0	0	1
7	1.90	1.20	3	0	3	0	2	0	2	0	2	0	2	0	2	0	1	0	3	0	0	1
8	1.90	1.20	3	0	3	0	2	0	2	0	2	0	2	0	2	0	1	0	3	0	0	1
9	1.70	1.16	3	0	2	0	2	0	2	0	2	0	2	0	1	0	1	0	3	0	0	1
10	1.70	1.16	3	0	2	0	2	0	2	0	2	0	2	0	1	0	1	0	3	0	0	1
11	-0.10	2.92	3	0	0	3	3	0	0	3	3	0	0	3	3	0	1	0	0	3	0	2
12	0.80	1.81	2	0	2	0	2	0	2	0	2	0	1	0	1	0	1	0	0	3	0	2
13	0.70	1.77	2	0	2	0	2	0	2	0	1	0	1	0	1	0	1	0	0	3	0	2
14	0.70	1.77	2	0	2	0	2	0	2	0	1	0	1	0	1	0	1	0	0	3	0	2
15	0.60	1.71	2	0	2	0	2	0	1	0	1	0	1	0	1	0	1	0	0	3	0	2
16	0.60	1.71	2	0	2	0	2	0	1	0	1	0	1	0	1	0	1	0	0	3	0	2
17	0.50	1.65	2	0	2	0	1	0	1	0	1	0	1	0	1	0	1	0	0	3	0	2
18	0.50	1.65	2	0	2	0	1	0	1	0	1	0	1	0	1	0	1	0	0	3	0	2
19	0.30	1.57	2	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	0	3	0	2
20	0.30	1.57	2	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	0	3	0	2
21	0.60	1.17	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	1	3	0	0	1
22	0.60	1.17	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	1	3	0	0	1
23	0.50	1.18	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	3	0	0	1
24	0.50	1.18	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	3	0	0	1
25	0.40	1.17	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	3	0	0	1
26	0.40	1.17	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	3	0	0	1
27	0.30	1.16	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	1
28	0.30	1.16	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	1
29	0.50	1.43	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	3	0	0	1
30	0.50	1.43	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	3	0	0	1
31	0.20	1.87	0	0	0	0	0	0	0	0	0	0	3	0	3	0	1	0	0	3	0	2
32	0.20	1.87	0	0	0	0	0	0	0	0	0	0	3	0	3	0	1	0	0	3	0	2
33	0.50	2.07	0	0	0	0	0	0	0	0	3	0	3	0	3	0	1	0	0	3	0	2
34	0.50	2.07	0	0	0	0	0	0	0	0	3	0	3	0	3	0	1	0	0	3	0	2
35	0.80	2.20	0	0	0	0	0	0	3	0	3	0	3	0	3	0	1	0	0	3	0	2
36	0.80	2.20	0	0	0	0	0	0	3	0	3	0	3	0	3	0	1	0	0	3	0	2
37	1.10	2.28	0	0	0	0	3	0	3	0	3	0	3	0	3	0	1	0	0	3	0	2
38	1.10	2.28	0	0	0	0	3	0	3	0	3	0	3	0	3	0	1	0	0	3	0	2
39	1.30	2.26	0	0	3	0	3	0	3	0	3	0	3	0	2	0	1	0	0	3	0	2
40	1.30	2.26	0	0	3	0	3	0	3	0	3	0	3	0	2	0	1	0	0	3	0	2

FIGURE 5-9, CODEC EVALUATION PRINTOUT

CODEC EVALUATION CALCULATION

SEQ	MEAN	STDEV	1	2	3	4	5	6	7	8	9	10
			EVALUATORS									
1	2.2	1.32	3	3	3	3	3	2	2	1	3	-1
2	2.2	1.32	3	3	3	3	3	2	2	1	3	-1
3	2.1	1.29	3	3	3	3	2	2	2	1	3	-1
4	2.1	1.29	3	3	3	3	2	2	2	1	3	-1
5	2	1.25	3	3	3	2	2	2	2	1	3	-1
6	2	1.25	3	3	3	2	2	2	2	1	3	-1
7	1.9	1.2	3	3	2	2	2	2	2	1	3	-1
8	1.9	1.2	3	3	2	2	2	2	2	1	3	-1
9	1.7	1.16	3	2	2	2	2	2	1	1	3	-1
10	1.7	1.16	3	2	2	2	2	2	1	1	3	-1
11	-0.1	2.92	[illegible]	-3	3	-3	3	-3	3	1	-3	-2
12	0.8	1.81	2	2	2	2	2	1	1	1	-3	-2
13	0.7	1.77	2	2	2	2	1	1	1	1	-3	-2
14	0.7	1.77	2	2	2	2	1	1	1	1	-3	-2
15	0.6	1.71	2	2	2	1	1	1	1	1	-3	-2
16	0.6	1.71	2	2	2	1	1	1	1	1	-3	-2
17	0.5	1.65	2	2	1	1	1	1	1	1	-3	-2
18	0.5	1.65	2	2	1	1	1	1	1	1	-3	-2
19	0.3	1.57	2	1	1	1	1	1	0	1	-3	-2
20	0.3	1.57	2	1	1	1	1	1	0	1	-3	-2
21	0.6	1.17	1	1	1	1	1	0	0	-1	3	-1
22	0.6	1.17	1	1	1	1	1	0	0	-1	3	-1
23	0.5	1.18	1	1	1	1	0	0	0	-1	3	-1
24	0.5	1.18	1	1	1	1	0	0	0	-1	3	-1
25	0.4	1.17	1	1	1	0	0	0	0	-1	3	-1
26	0.4	1.17	1	1	1	0	0	0	0	-1	3	-1
27	0.3	1.16	1	1	0	0	0	0	0	-1	3	-1
28	0.3	1.16	1	1	0	0	0	0	0	-1	3	-1
29	0.5	1.43	1	0	0	0	0	0	3	-1	3	-1
30	0.5	1.43	1	0	0	0	0	0	3	-1	3	-1
31	0.2	1.87	0	0	0	0	0	3	3	1	-3	-2
32	0.2	1.87	0	0	0	0	0	3	3	1	-3	-2
33	0.5	2.07	0	0	0	0	3	3	3	1	-3	-2
34	0.5	2.07	0	0	0	0	3	3	3	1	-3	-2
35	0.8	2.2	0	0	0	3	3	3	3	1	-3	-2
36	0.8	2.2	0	0	0	3	3	3	3	1	-3	-2
37	1.1	2.28	0	0	3	3	3	3	3	1	-3	-2
38	1.1	2.28	0	0	3	3	3	3	3	1	-3	-2
39	1.3	2.26	0	3	3	3	3	3	2	1	-3	-2
40	1.3	2.26	0	3	3	3	3	3	2	1	-3	-2
1.2	MEAN		1.5	1.4	1.5	1.4	1.5	1.4	1.6	0.5	0.0	-1.5
	STD. DEV.		1.2	1.3	1.2	1.3	1.2	1.3	1.2	0.9	3.0	0.5

FIGURE 5-10, CODEC EVALUATION CALCULATION EXAMPLE

of the method. Most of the data has been randomized to the point where it is felt that it is valid for descriptive purposes. However, some data entries were forced in order that certain unusual output conditions would be demonstrated.

For sequence 11, the data entry produced an unusually high standard deviation. An examination of the data entered for sequence 11 shows that the evaluators disagreed rather strenuously in their evaluation of the performance of codec A as compared to codec B. The grades alternate between +3 and -3 for the most part. If this condition exists for other codec pair evaluations, an examination of the validity of that sequence would be in order. Note further the standard deviation for evaluator 9. Not only is the mean for this evaluator out of line with the other means and may therefore be suspect, the standard deviation value is also very high. Examination of the data shows that this evaluator's grades were totally inconsistent throughout and therefore may be suspect. Finally, the mean for evaluator 10 is inconsistent with that of any of the other evaluators indicating that this data should be reviewed.

Inconsistencies of this type are normally not anticipated. They are described here to show that precautions against eratic data have been included in the data processing effort.

SECTION 6 CONCLUSION

The testing methodology described in the preceding paragraphs provides a practical method for the evaluation and grading of digital 1.544 Mbps color television codecs capable of presenting motion. The tests consist of passing the video signal from a video tape recorder through codec transmission systems and recording the output signals. The input video test tape contains specially designed pictures and sequences to permit thorough evaluation of the performance of the codecs. The output video tapes are subjectively evaluated to determine relative performance grades.

The results of the test will provide an evaluation of the performance of each codec in the following areas.

- Motion Capability
- Output picture quality
- Effects of transmission channel errors

The codecs will be ranked in order of comparative performance using the criteria listed above. The relative weighting of these criteria depends on each evaluator and is part of his composite score.

Substantial additional raw data will be provided to permit expanded future evaluation for various purposes. Video test signals will be passed through the codecs and the output recorded for future evaluation. The input video test tape can also be compared to the codec output tapes at some time in the future to

generate an impairment grade should that be desireable.

The results of the evaluation will provide truly meaningful data for the preparation of performance and test specifications for codecs of this type.

APPENDIX A

OBJECTIVE MOTION EVALUATION TEST

The motion performance of television systems has traditionally been evaluated on a subjective rather than an objective basis. This has been adequate for a number of reasons. The primary factor is that the transmission is in real time and therefore does not affect the motion characteristics in any measure worth considering. Secondly, the motion characteristics of the system have been masked to a large degree by the retention in the television camera tube and to a lesser degree by the human eye.

These conditions changed drastically with the advent of digital television systems operating at reduced data rates. The necessity for reduced data rates does not permit the transmission of video data in real time while preserving full resolution, gray scale, and color capability. Many very clever techniques have been developed and are used to minimize any observed degradation of picture quality due to the reduction in transmission data rate. Various equipment manufacturers have been very successful in employing these techniques to produce low data rate television systems providing satisfactory performance. Therefore, motion performance for a fixed resolution, gray scale, and color capability is now closely related to the transmission data rate. The masking influence of the television camera tube is no longer a major parameter in determining motion performance. The motion performance limitation of the channel due to the data rates used

in digital televisions systems of the type under consideration is a substantially more important factor than that introduced by the camera tube to the point that the latter can be ignored. Secondly, systems of this type often operate with electronically generated data which produce instantaneous changes. Therefore, motion must be considered on a new basis.

Subjective tests are an excellent method of determining the motion performance of a digital television codec. This is particularly true because the type of motion degradation is highly dependent on the signal processing algorithm employed in the codec. The methods of performing subjective tests are well understood and, in general, well accepted. However, they are very time consuming and require a large number of evaluators if many of tests are to be made (typically 10 evaluators per test). Therefore, it is very desirable to develop an objective, quantitative motion performance parameter which is truly meaningful and which can be applied easily to any codec to produce consistent unambiguous data.

The technique incorporated into the codec signal processing algorithm to effect the transmission of motion is most commonly a method of transmitting on a priority basis the change information between successive frames (which in effect constitutes the motion data) over a period of time which substantially exceeds one real time television frame interval (1/30 sec.). As a result, a considerably longer period of time is required for the change information to completely update the picture. The duration of this period is determined by the magnitude of the change, the data

rate, and the signal processing algorithm. The common element is time. Therefore it appears that time is an excellent indicator of the motion performance of a codec; specifically, it is the time required from the occurrence of the change in the input picture until all of the change information defining the motion has been incorporated and the output picture has become stable. In effect, this is a measure of the time from stimulus to completion of response.

Time as a motion performance parameter is very meaningful in that the evaluator can relate to it quite readily in terms of its physical significance. In addition, it is an extremely flexible parameter in that it can be applied to the measure of many different types of motion (changes in picture content). For example, the change may be from a black to a white field, a blue to a yellow field, one set of alpha-numeric data to another, one scene to another, etc. It could result from camera panning or zooming, motion in the scene, or switcher punching, mixing, wiping, or keying between two video signals.

An approach for an unambiguous and meaningful objective measurement of codec motion performance is described below.

The concept consists of determining precisely the time interval required for a codec system to completely update the display produced by its output signal after a change in the transmitted picture content has occurred. Specifically, this is the time interval between the initiation of the change in the picture and its completion. Two precisely definable television

signals are applied to a video switch operating during the vertical blanking interval between fields. The video switch permits the selection of which of the two signals is applied to the transmit codec of the codec system. The output of the receive codec is applied to a television monitor so that the process may be observed by the evaluator. The test consists of the following process. Although any pair of television signals can be used as the input test signals the following are postulated for purposes of description. Video "A" is a black field with a white square in its center occupying about 25% of the picture area. Video "B" is a totally black field. A possible alternate is yellow and blue fields which result in a big step in luminance and chrominance levels and a 180° chrominance phase shift. When the test is initiated, Video "A" is being transmitted and the output signal will provide a stable presentation of, in this example, a white square set in a black background. When the video switch is operated, the white square signal is replaced by an all black signal. Transmission of the all black field will commence. The display of the output signal will begin to change and after some period of time will again produce a stable display, this time of the all black signal. The interval of time during which the display is changing is a measure of the motion capability of the codec system. The concept is quite simple and flexible. Selecting the appropriate test signals will permit the evaluation of the contribution of various parameters to the motion performance; eg., luminance, chrominance, etc. Experimentation to select the best pair of test signals is still required.

The objective motion evaluation test described above can be implemented rather simply. The unit of time must of necessity be a field interval because the stimulus provided to the user is the output display which is updated on a field basis. The test process consists of determining how many field intervals have elapsed from the initiation until the completion of the change.

A simple method of implementing the elapsed time count is to record the output video signal from the receive codec on a video tape recorder with a single frame playback capability. The number of fields can then be precisely determined by viewing the display from the video tape on picture and waveform monitors and advancing the tape on a frame by frame basis. The number of fields during which change in the picture is evident are counted and produce a figure of merit for that particular motion codec.

A somewhat more sophisticated approach can be taken to completely eliminate participation by an evaluator. The specific signals selected as the test signals in the above sample permit the use of conventional test equipment to determine the time interval required for the codec to complete the display update. The initial transmitted signal consists of a white square on a black background occupying about 25% of the picture area. A counter connected to the output of the receive codec can count the number of lines on which white data exists. This requires only that the video signal be clamped, and that the counter can be adjusted to read slightly above black level and have adequate speed capability. It is required that the counter be capable of printing out or storing its cumulative count in response to an

external command. The requirements listed are not sophisticated. The vertical drive signal is used to trigger the counter to end the accumulation period and store or print out the data for a field interval. During the period when the white square is being transmitted the count will be consistent from field to field. As soon as the switcher is exercised, the codec output will begin to change and the recorded count will change correspondingly. When the change has been completed, that is, an all black signal is displayed, the count during any field interval will be zero. Examining the printout or the stored data will permit precise determination of the number of field intervals required to complete the change. In this way a totally objective figure of merit can be determined.

The test signals described above have been incorporated on the test video tape so that the described procedure can be evaluated. To date such tests have not yet been performed. The signals, white square and black signals and the yellow and blue signals, initially appear to be well suited for this purpose. They may need to be optimized based on experience with actual codec testing.

One word of caution is required. The test signals incorporated in the test tape produce drastic changes of many or all picture elements but the pictures both before and after the change are extremely simple. Depending on the algorithm of the codec, changes between such pictures may be easier to reproduce than many ordinary types of motion produced by camera panning, zooming, or motion of subjects because the same updating process

is used for large areas and possibly the whole picture. Therefore, these switched signals may not constitute a sufficiently stringent test, and performance with them may not be correlated with conventional motion performance. Should early tests indicate that this is the case, further effort will be needed to establish more complex electronically generated signals which more closely simulate actual motion performance. Devices are available to produce an essentially unlimited number of well definable artificial signals which lend themselves for convenient objective evaluation.

END

FILMED

12-85

DTIC